

Алексей Шпагин



Алексей Шпагин

- Руководитель команды бэкенда ВК Звонки
- 10 лет работы в VoIP-телефонии и видеозвонках
- Бэкграунд — разработчик C++
- В руководстве командами 4 года



VK Звонки

Бесплатные звонки без ограничений по времени и количеству участников.



Для работы и учёбы

Демонстрация экрана в 4K, трансляция, планирование и запись.



Управление звонками

Зал ожидания, управление микрофонами, функция «Поднять руку» и другие возможности модерации.



Технологичность

Интеллектуальное шумоподавление, собственная AR-технология замены фона.

20 млн

пользователей общаются в VK Звонках ежемесячно

Источник: ВКонтакте, апрель 2022 г.



6 млн

звонков в день

20 млн

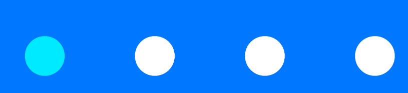
пользователей в месяц

15 тыс.

одновременных звонков

Содержание

Из чего складывается качество звука?



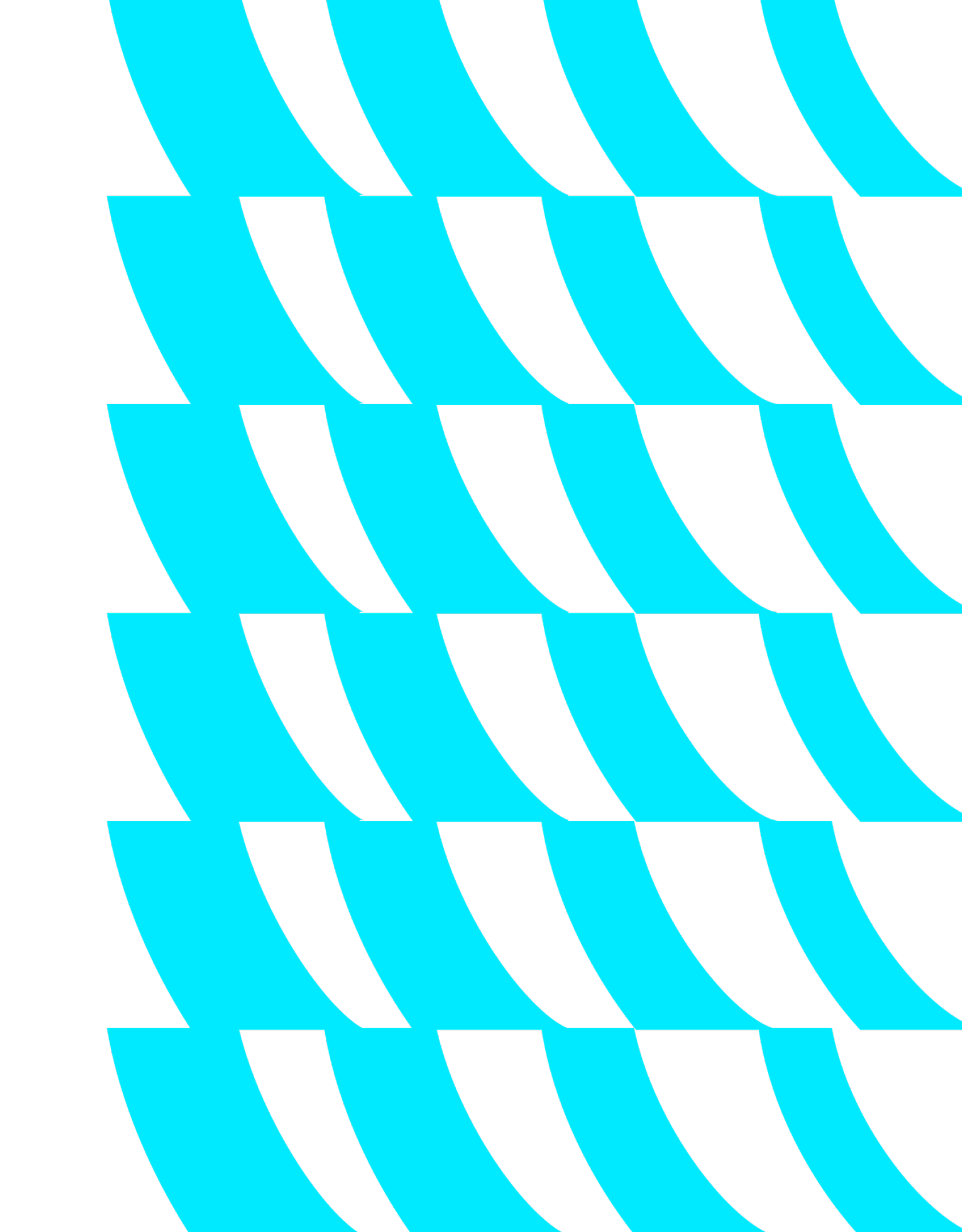


Видеозвонки — замена живым встречам

- При встрече офлайн все отлично друг друга видят и слышат
- Система видеозвонков вносит искажения в передачу звука
- Если искажения сильные, общаться некомфортно, и встреча перестает быть похожа на встречу живьем

Требования к передаче звука в системе видеозвонков

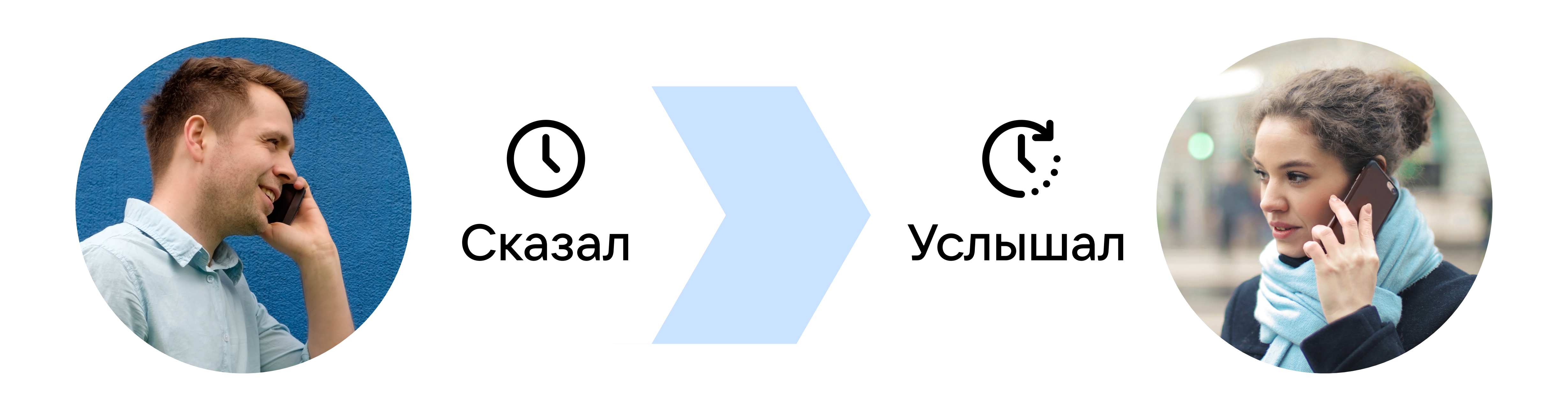
В порядке убывания критичности





Непрерывность звукового потока

Минимальная latency



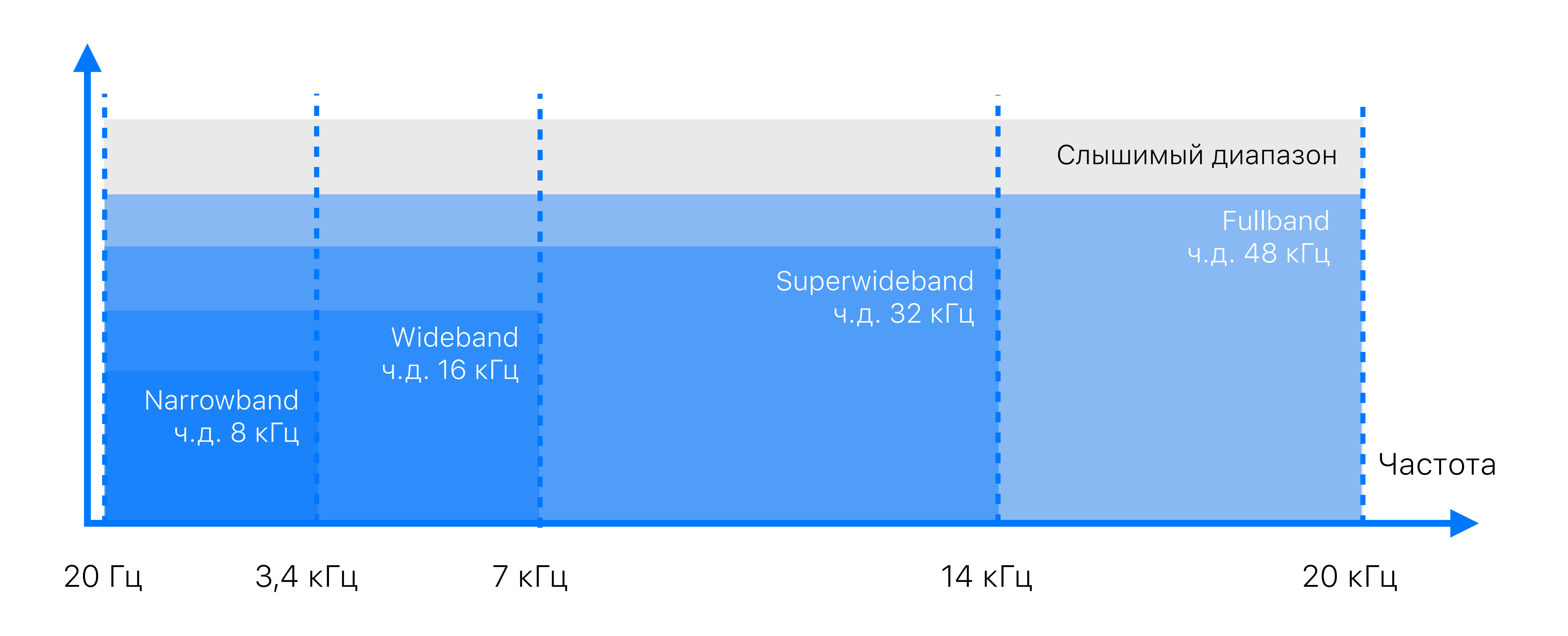
Задержка



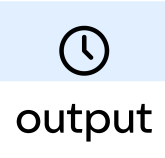
Отсутствие артефактов в звуке

- Постоянный треск
- Периодические щелчки в произвольное время
- Клиппинг

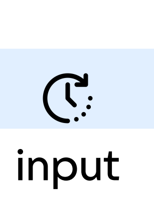
Достаточный частотный диапазон







output



input







Измерение качества звука





Что же будем оценивать?

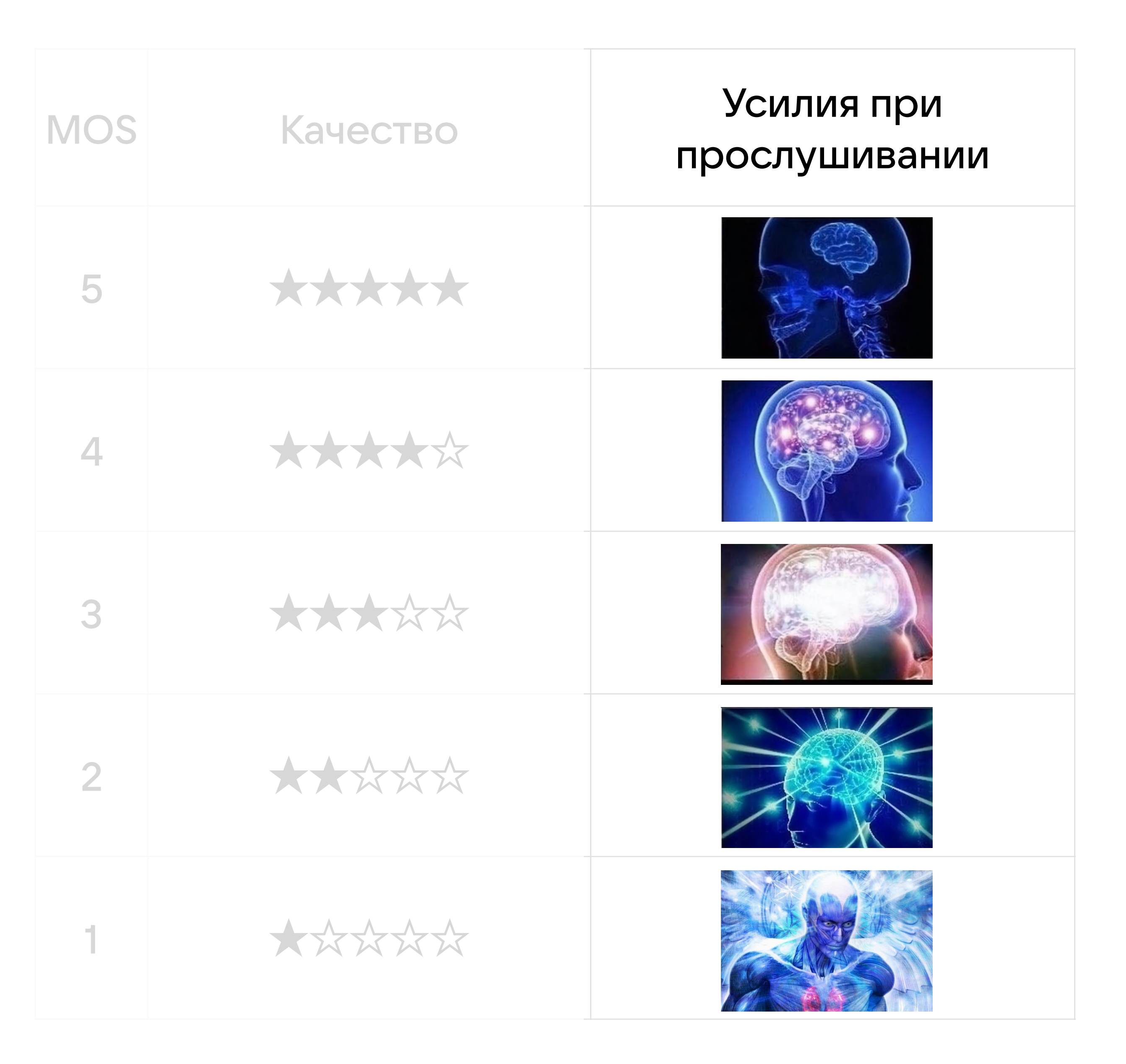
Не звук абстрактно и не качество звукозаписи — а речь:

- Степень искажений
- Разборчивость
- Комфортность восприятия

MOS - Mean Opinion Score

MOS	Качество	Усилия при прослушивании
5	★★★★★	
4	★★★★☆	
3	★★★☆☆	
2	★★☆☆☆	
1	★☆☆☆☆	

MOS - Mean Opinion Score

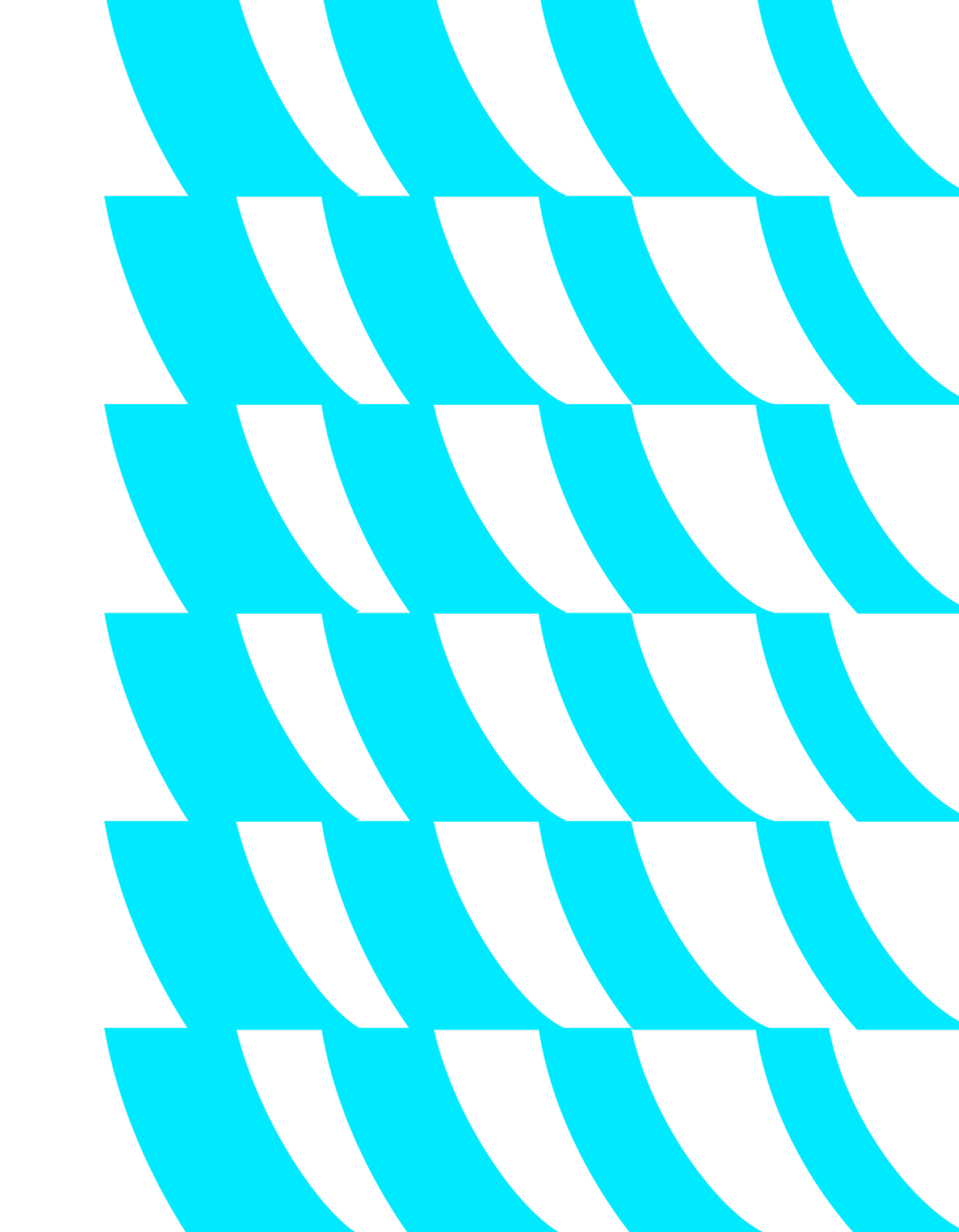


MOS	Качество	Усилия при прослушивании
5	★★★★★	
4	★★★★☆	
3	★★★☆☆	
2	★★☆☆☆	
1	★☆☆☆☆	

MOS, равный 5 — недостижим

Кодек	**Max. MOS**
G.711	4,4
G.722	4,5
G.729	3,92
OPUS*	4,5

Методики и алгоритмы измерения качества речи

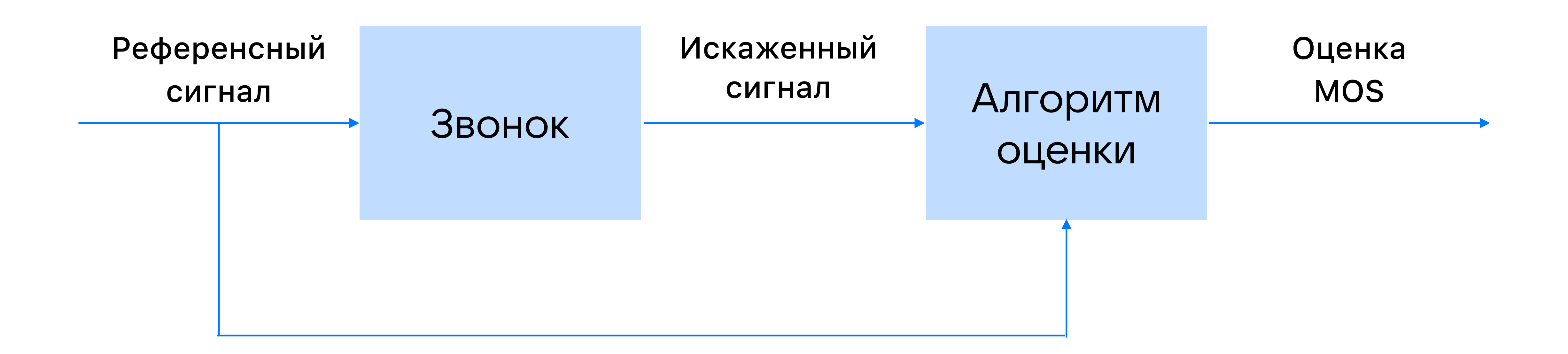




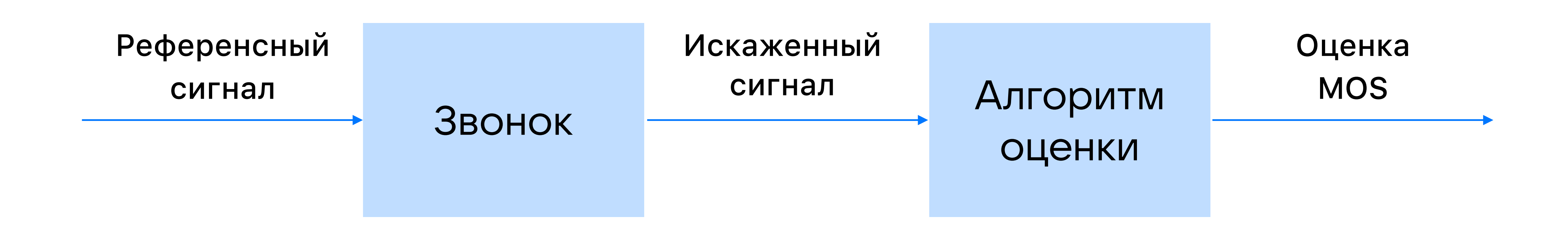
Ориентируемся на восприятие речи человеком

- Алгоритмы измерения качества речи предсказывают, как бы живой человек оценил заданный звуковой фрагмент
- Оценка MOS показывает то, как человек воспринимает фрагмент речи

Методика оценки качества речи на основе референса



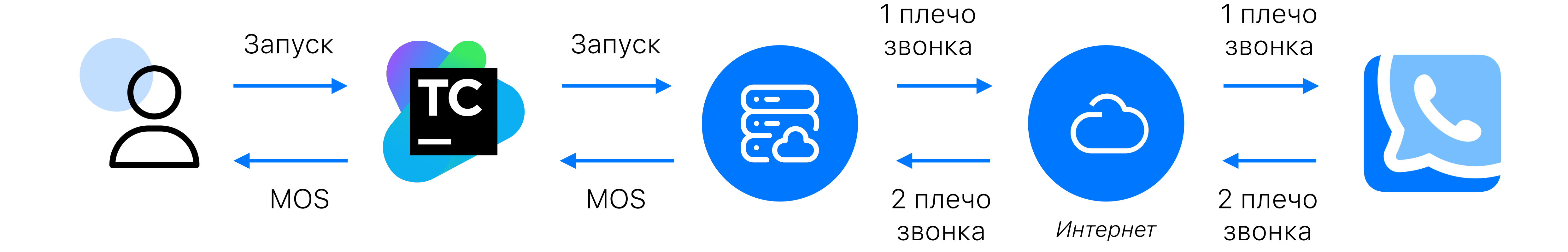
Безреференсная методика оценки качества речи



Методики оценки качества речи

Мы пользуемся в VK Звонках	Мы не пользуемся
ViSQOL Open source аналог POLQA	**POLQA** Коммерческое решение
NISQA Безреференсная методика. Open source	**AQuA** Перспективное решение, но пока проигрывает POLQA
	PESQ Устаревшее коммерческое решение

Стенд для измерения качества голоса в VK Звонках



Итого про оценку качества речи

1. Измеряем качество голоса, речи
2. Используем метрику MOS от 1 до 5.
3. MOS не бывает равен 5
4. Предсказание реакции человека на звуковой фрагмент
5. Референсные и безреференсные подходы

Примеры проблем при передаче звука, которые мы решали



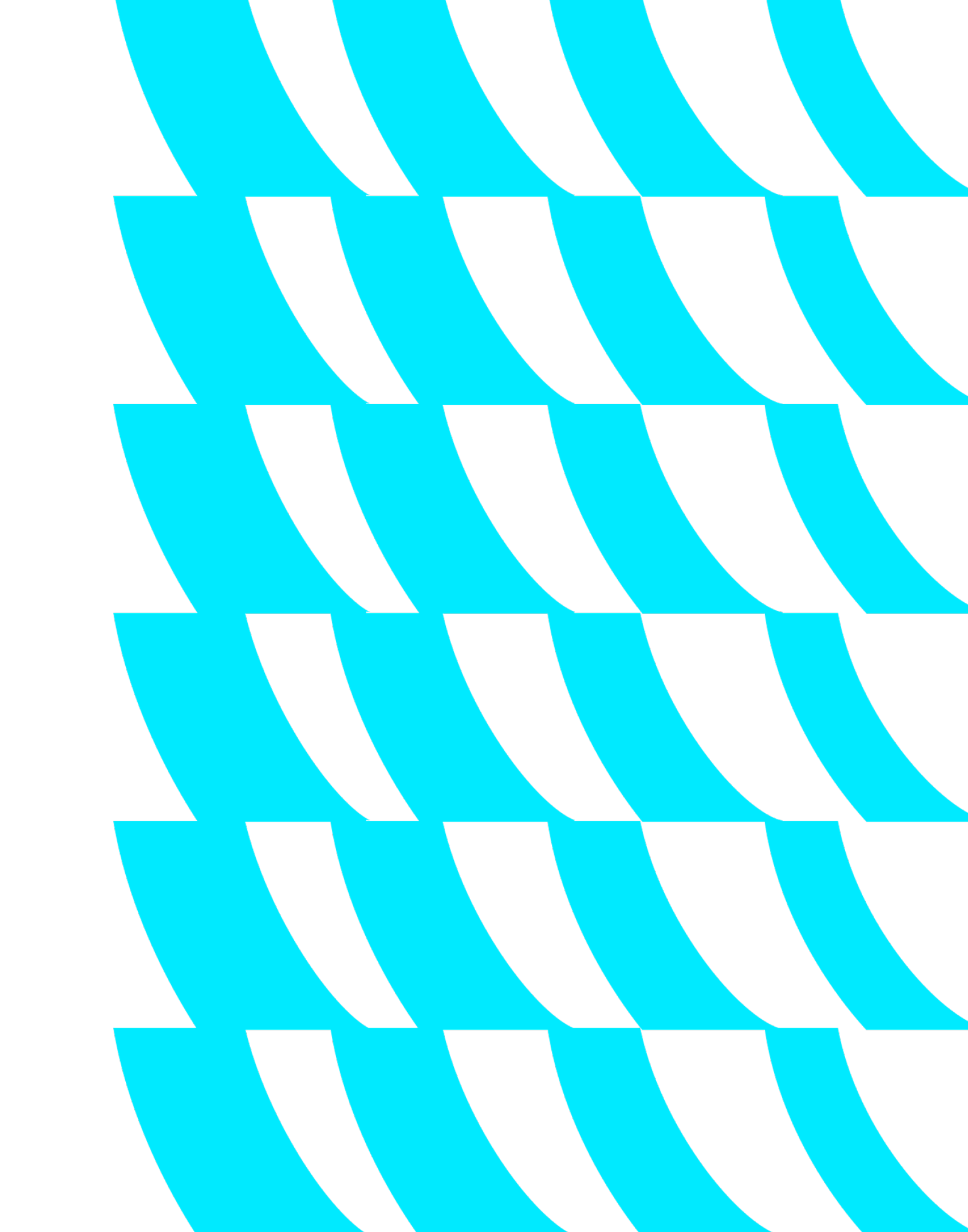
Виды проблем при передаче голоса через Интернет

- Проблемы, обусловленные особенностью передачи данных по TCP/IP-сетям
- Проблемы акустического характера

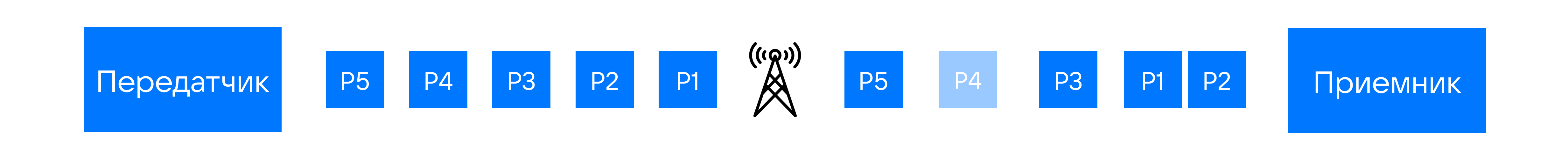




Проблемы, обусловленные особенностью передачи данных по сети



Принцип передачи голоса по сети



Пакеты могут:

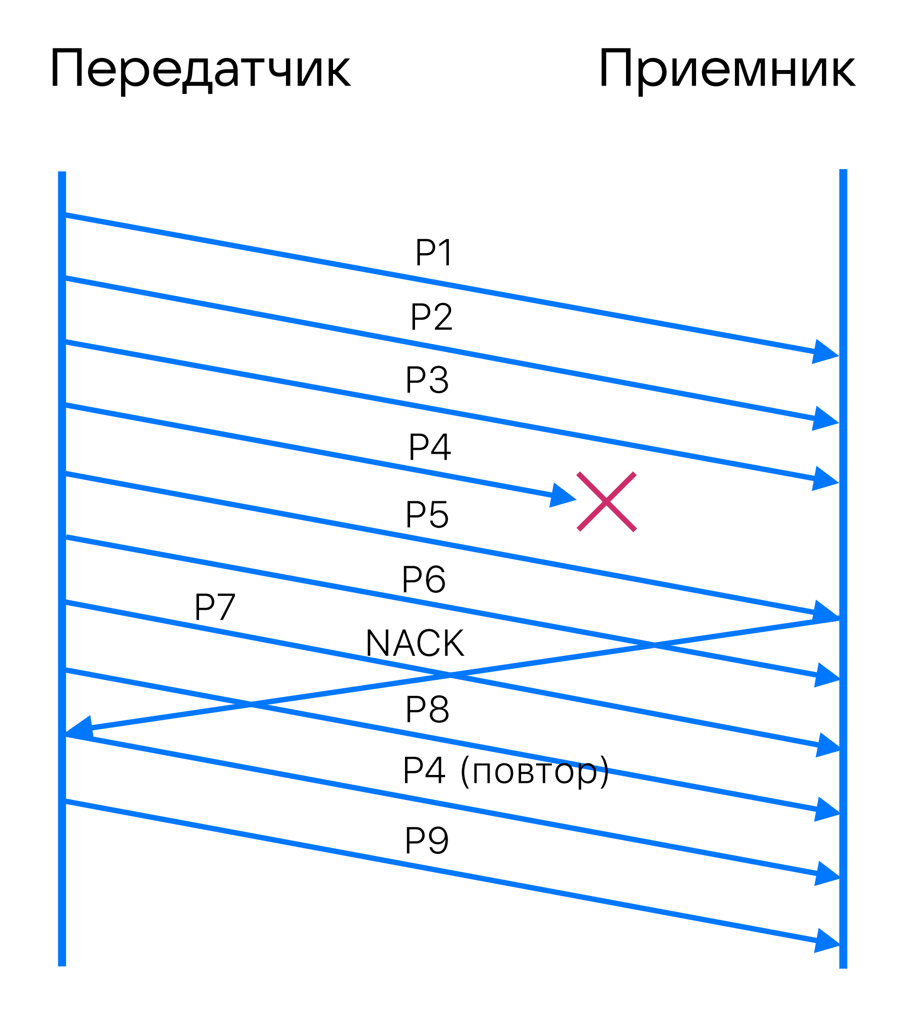
- теряться
- «дрожать» (Jitter)
- задерживаться (delay)
- меняться местами (reordering)

*Проблемы проявляются сильнее на мобильном интернете, при проводном подключении - слабее

Jitter Buffer

- Компенсирует «дрожание»
- Выравнивает трафик
- Может «подождать» недошедший вовремя пакет
- Больше Jitter Buffer — больше latency
- В среднем ~100 — 200мс

NACK — Negative Acknowledgment



- Позволяет перезапросить потерянный пакет
- Хорошо работает на коротких RTT
- При длинных RTT не имеет смысла

FEC — Forward Error Correction

- Кодек добавляет в битстрим дополнительную информацию для восстановления
- Хорошо работает на больших RTT

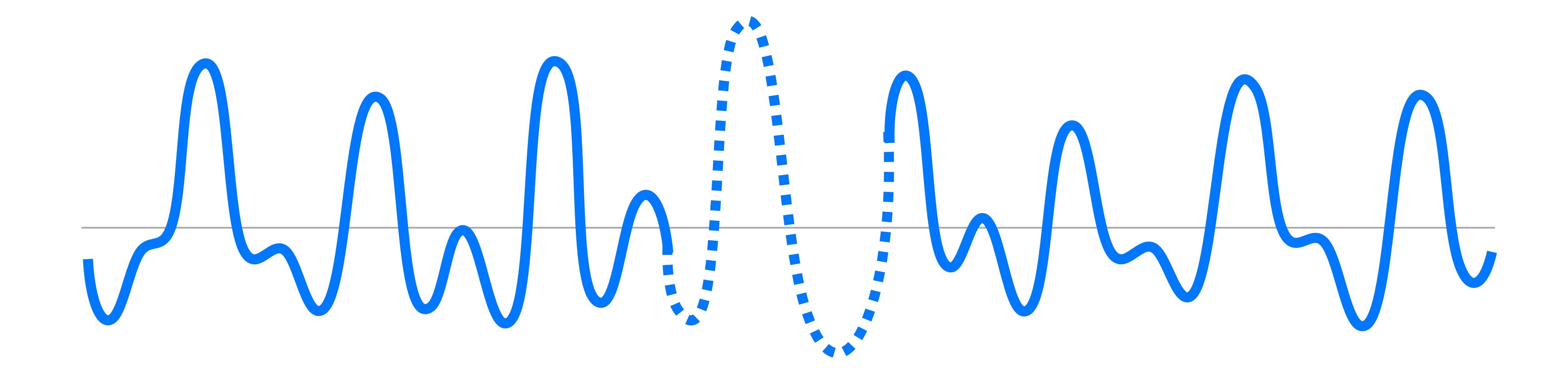
Условия	**Средний MOS**
loss = 5%, FEC = false	3,94
loss = 5%, FEC = **true**	**4,19**

RED — Redundancy (RTP Payload for Redundant Audio Data)

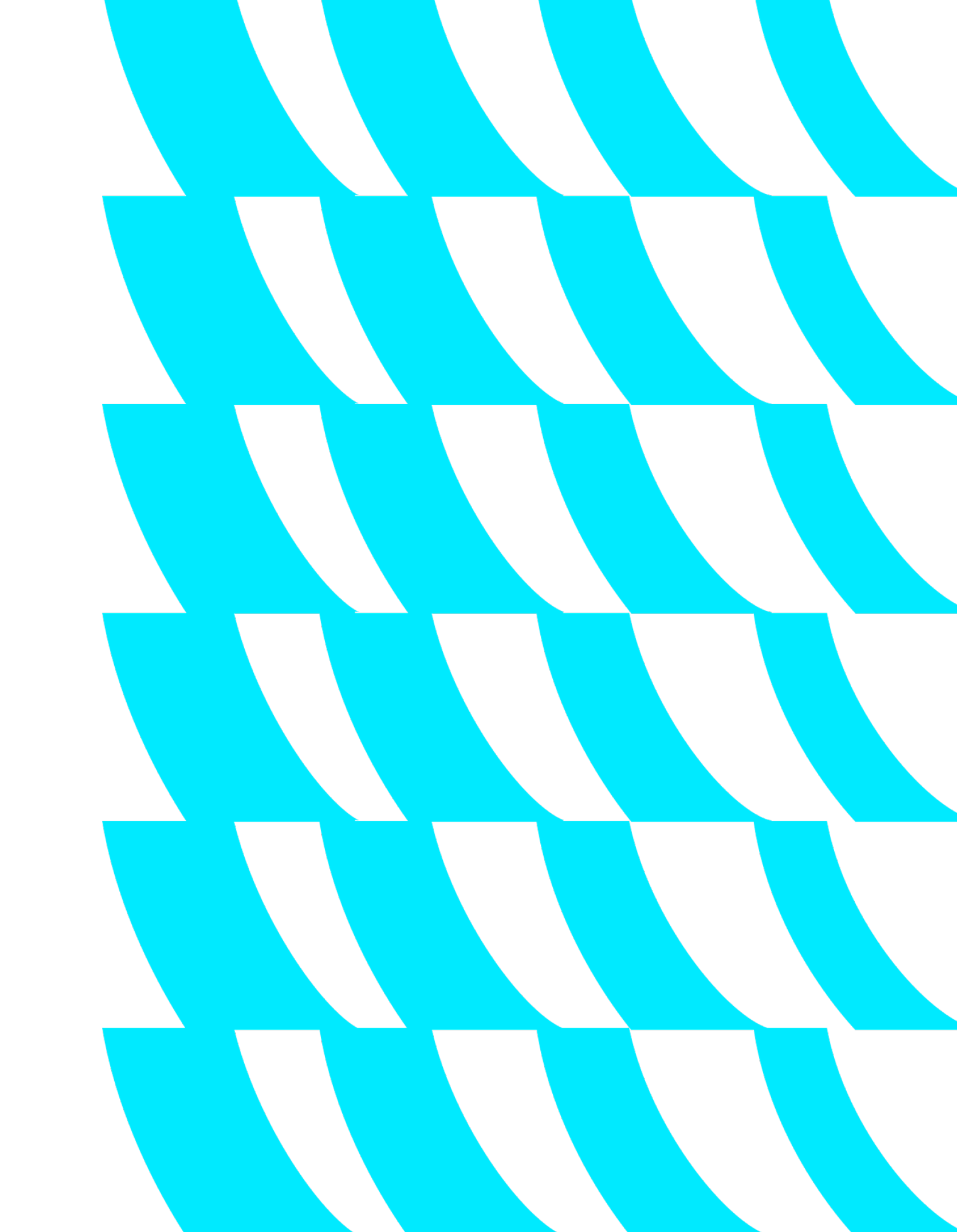
- Избыточность на уровне RTP-пакетов
- В один RTP-пакет помещается 2 или более аудиокадров
- Разница с FEC: избыточная информация не в битстриме, а объединяется несколько битстримов

Условия	**Средний MOS**
loss = 10%, delay = 200 ms, RED = false	3,60
loss = 10%, delay = 200 ms, RED = **true**	**4,14**

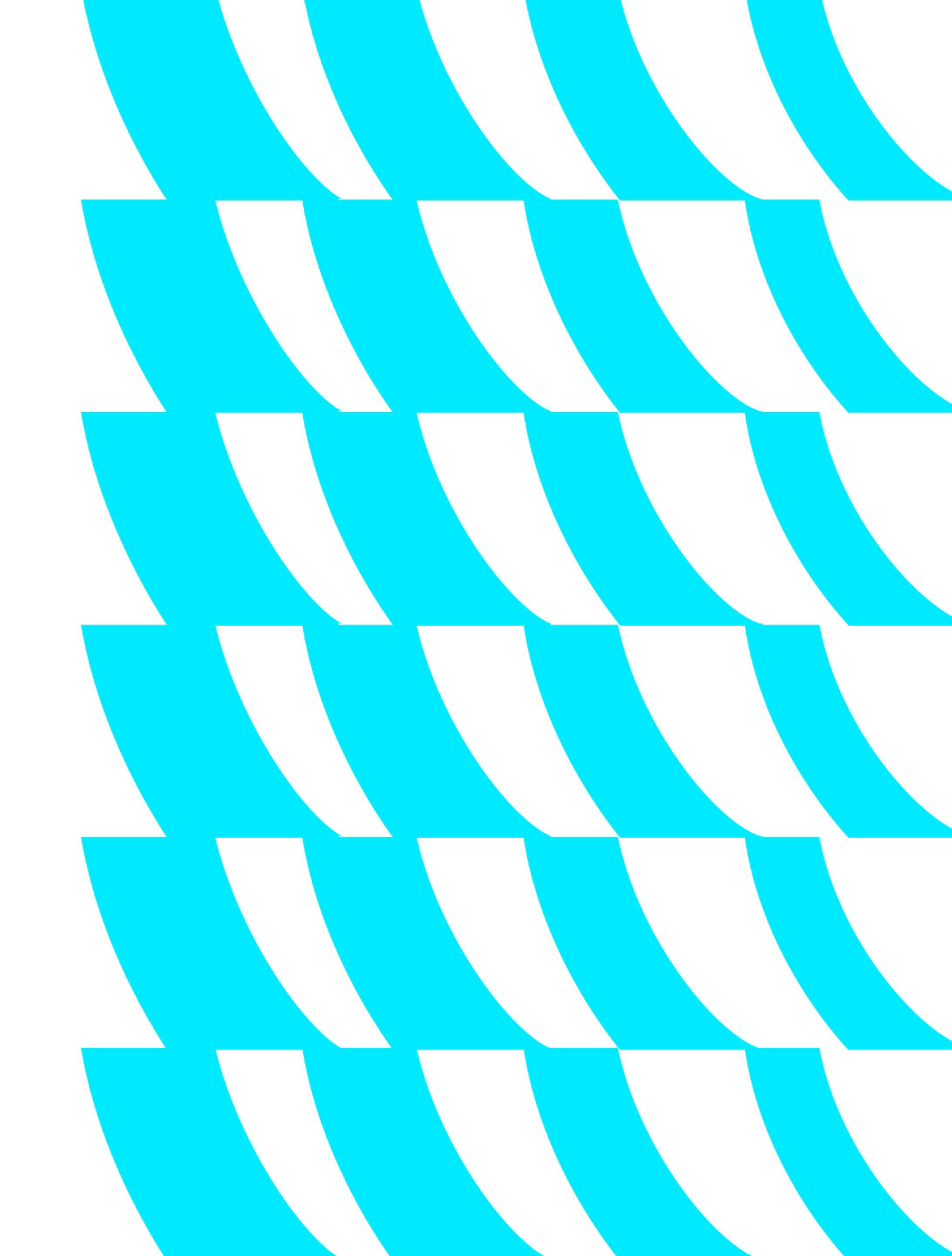
PLC — Packet Loss Concealment



Проблемы акустического характера



Внешние шумы

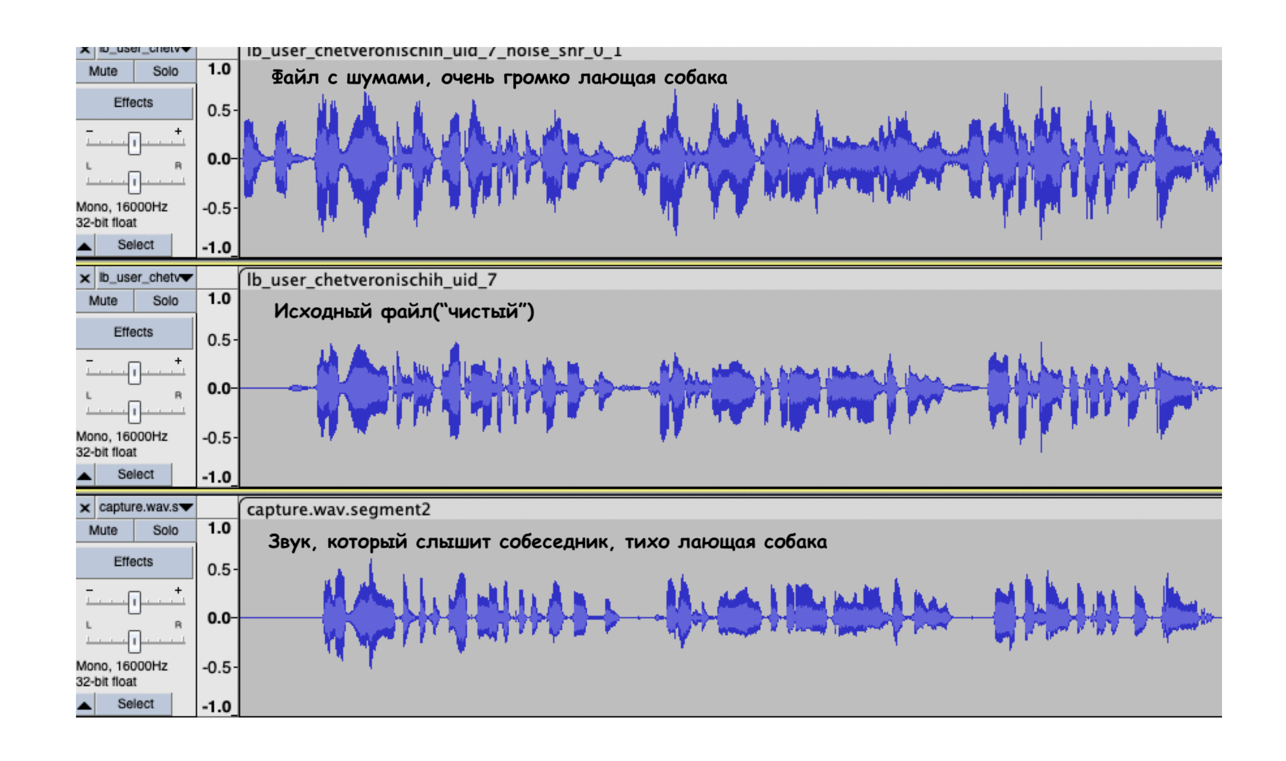


Источники шума



Шумоподавление





Условия	Средний MOS
Референс без шумов	3,91
Референс + шумы без шумодава	~ 1,00
Референс + шумы + шумоподавление	2,79

VAD — Voice Activity Detection

- Вставляет тишину там, где нет голоса
- Собственное решение на базе ML

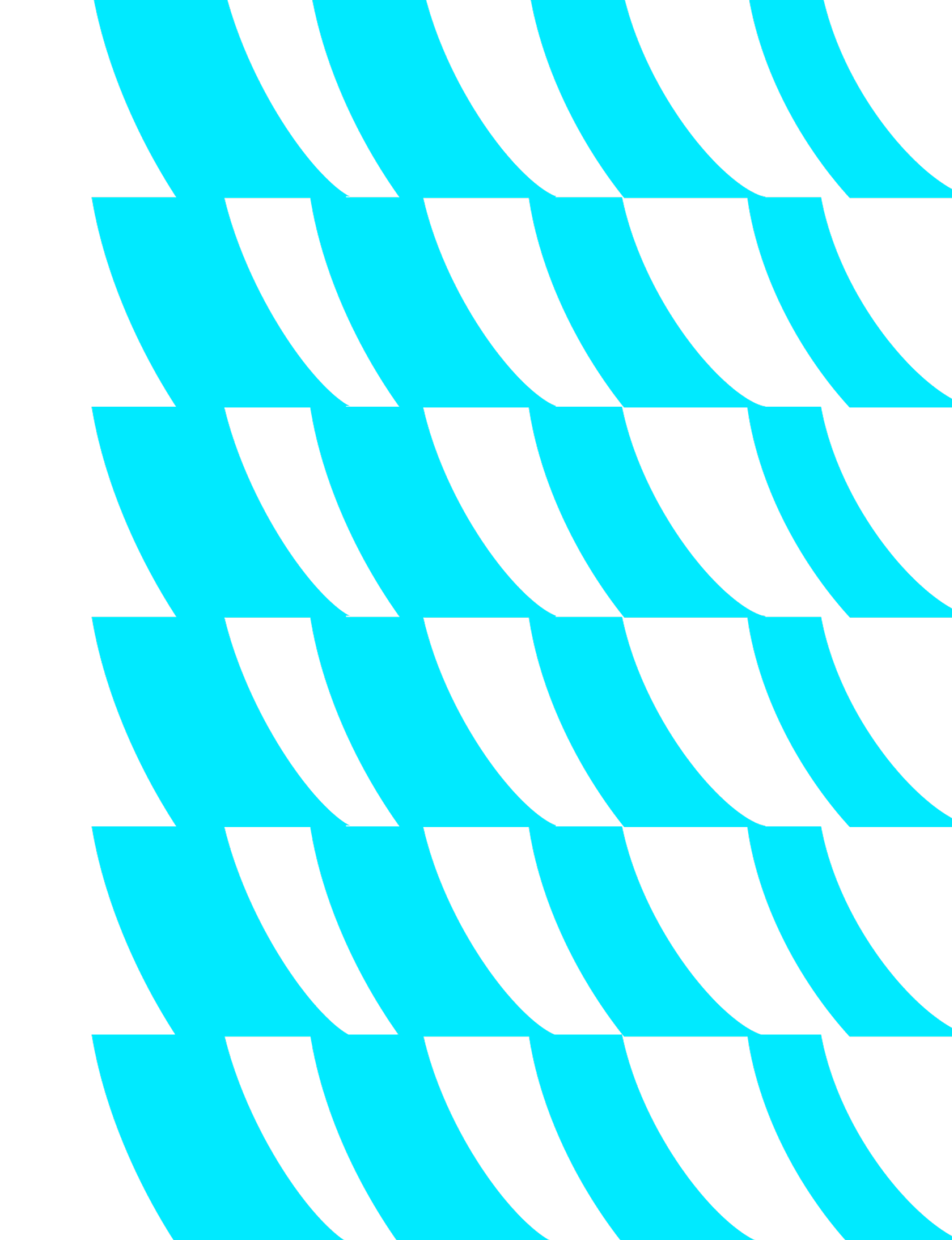
Условия	**Средний MOS**
VAD выключен	3,81
VAD включен	**3,67**

SNR — Signal to Noise Ratio

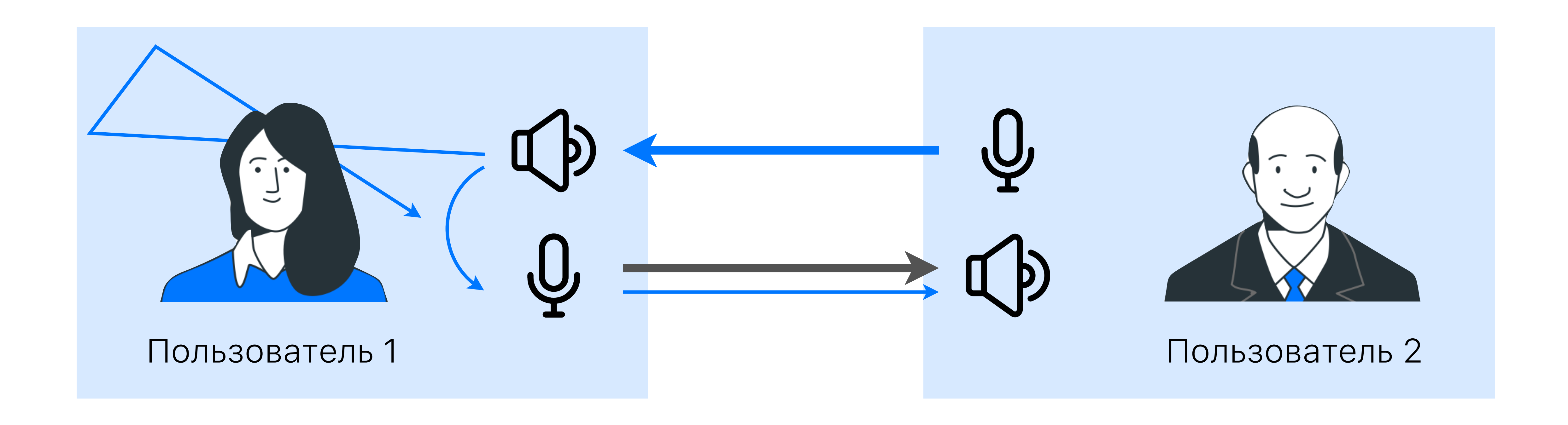
- Определяет, присутствует ли шум в звуке от участника звонка
- Если есть шум, то включаем шумодав, если нет, то не включаем, чтобы сохранить качество голоса
- Учимся оценивать качество работы SNR

Условия	**Средний MOS**
VAD — включен SNR — выключен	3,67
VAD и SNR - оба включены	**3,46**

Акустическое эхо



Возникновение эха



Эхоподавление

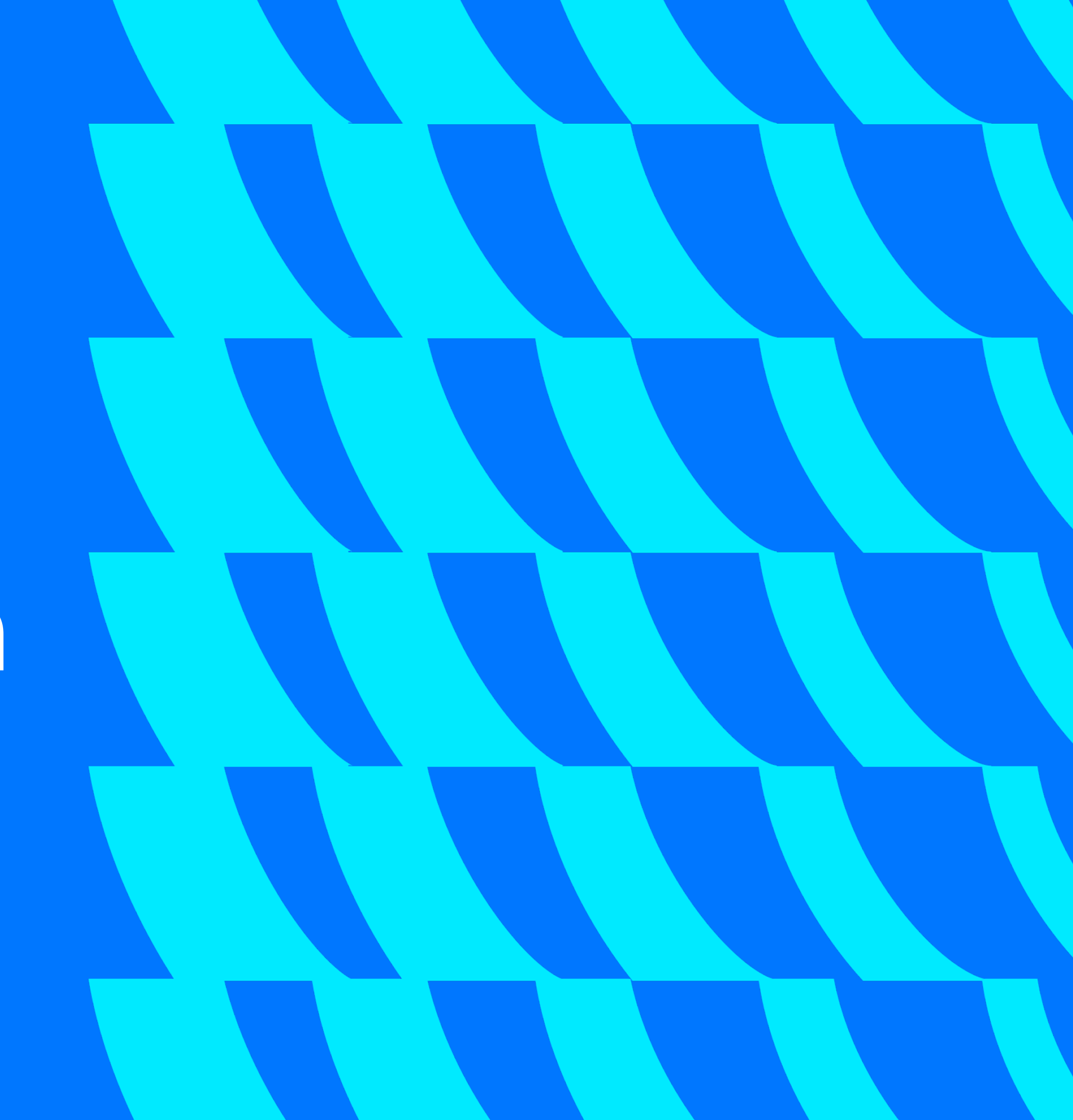
- Используем эходав WebRTC
- Снижаем громкость динамиков, если пользователь начинает говорить
- Проблема double talk

Проблемы, с которыми мы сталкиваемся

1. Шумоподавление
2. VAD
3. SNR - Signal to Noise Ratio
4. Эхоподавление

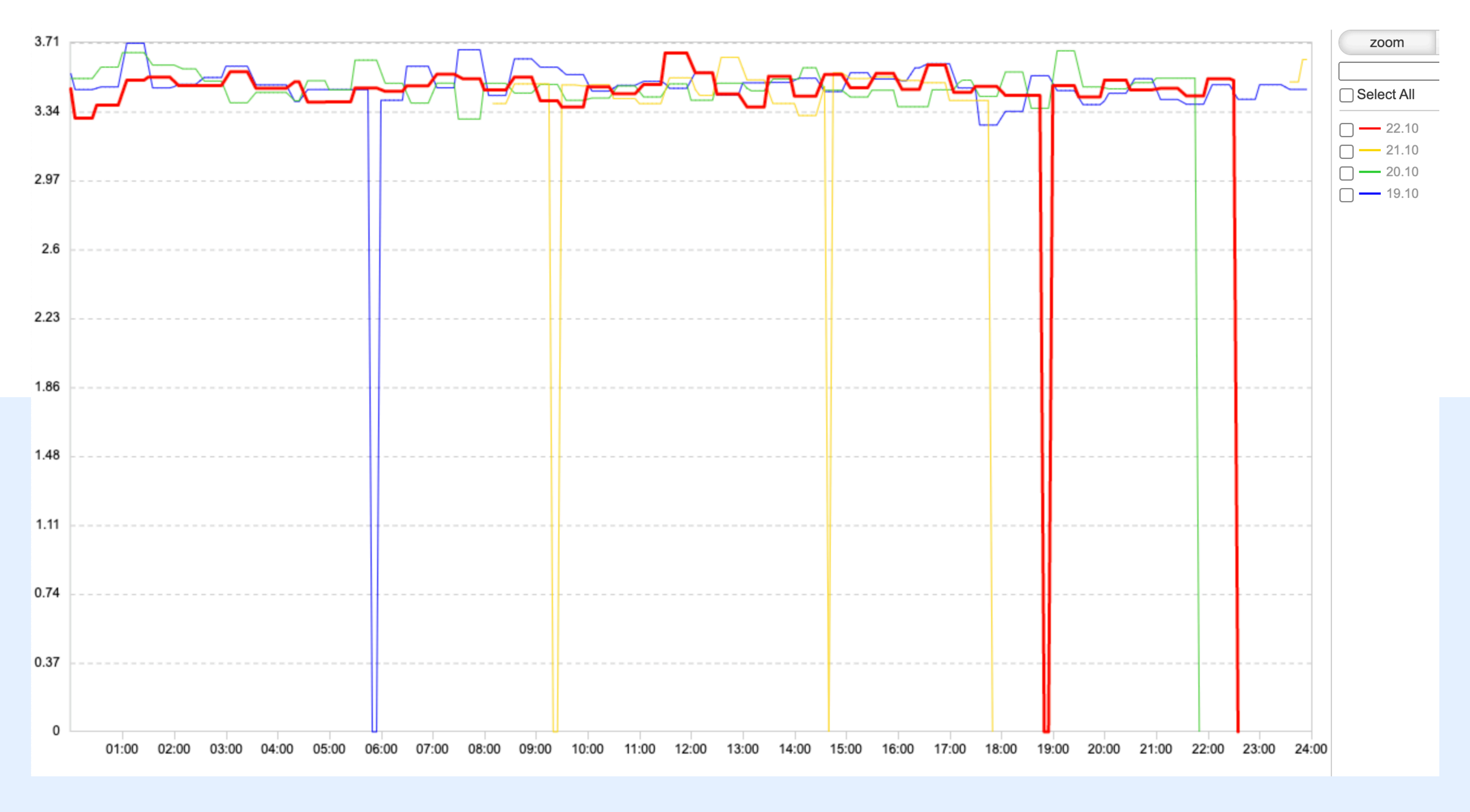
4. Jitter Buffer
5. PLC — Packet Loss Concealment
6. NACK — Negative Acknowledgment
7. FEC — Forward Error Correction
8. RED — Redundancy

Применение инструментов измерения качества голоса



Варианты использования джобы для оценки качества

- На dev-окружении при разработке фичи
- На регрессии перед релизом
- Сравниваем разные версии продукта
- Оцениваем эффект от новой фичи
- Мониторим продакшн



Мониторинг качества голоса на продакшене

Обратная связь и комментарии по докладу по ссылке

Алексей Шпагин,
ВКонтакте

http://vk.com/adshpagin
http://t.me/adshpagin



